



Harm–benefit analysis – what is the added value?

A review of alternative strategies for weighing harms and benefits as part of the assessment of animal research

Grimm, Herwig; Olsson, I. Anna S.; Sandøe, Peter

Published in:
Laboratory Animals

DOI:
[10.1177/0023677218783004](https://doi.org/10.1177/0023677218783004)

Publication date:
2019

Document version
Peer reviewed version

Citation for published version (APA):

Grimm, H., Olsson, I. A. S., & Sandøe, P. (2019). Harm–benefit analysis – what is the added value? A review of alternative strategies for weighing harms and benefits as part of the assessment of animal research. *Laboratory Animals*, 53(1), 17-27. <https://doi.org/10.1177/0023677218783004>

Harm-Benefit Analysis – What is the added value?

A review of alternative strategies for weighing harms and benefits as part of the assessment of animal research[§]

Herwig Grimm^{1*}, I Anna S Olsson^{2*#}, Peter Sandøe³

¹Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University of Vienna and University of Vienna, Austria

²Laboratory Animal Science, IBMC – Instituto de Biologia Molecular e Celular, i3S - Universidade do Porto, Portugal

³Department of Food and Resource Economics and Department of Veterinary and Animal Sciences, University of Copenhagen, Denmark

* Herwig Grimm and Anna Olsson share the first authorship

Correspondence should be directed to Dr. Anna Olsson, Laboratory Animal Science, IBMC - Institute for Molecular and Cell Biology, i3S - Universidade do Porto, Rua Alfredo Allen 208, 4200-135 Porto, Portugal or Olsson@ibmc.up.pt

Abstract

Animal experiments are widely required to comply with the 3Rs, to minimise harm to the animals and to serve certain purposes in order to be ethically acceptable. Recently, however, there has been a drift towards adding a so-called harm-benefit analysis as an *additional* requirement in assessing experiments. According to this, an experiment should only be allowed if there is a positive balance when the expected harm is weighed against the expected benefits. This paper aims to assess the added value of this requirement. Two models, the *discourse* and the *metric* model, are presented. According to the former, the weighing of harms and benefits must be conducted by a committee in which different stakeholders engage in a dialogue. Research into how this works in practice, however, shows that in the absence of an explicit and clearly defined methodology, there are issues about transparency, consistency and fairness. According to the metric model, on the other hand, several dimensions of harms and benefits are defined beforehand and integrated in an explicit weighing scheme. This model, however, has the problem that it makes no real room for ethical deliberation of the sort committees undertake, and it has therefore been criticised for being too technocratic. Also, it is unclear who is to be held accountable for built-in ethical assumptions. Ultimately, we argue that the two models are

[§] The reference of the original version is:

Grimm, H., Olsson, I. A. S., & Sandøe, P. (2019). Harm–benefit analysis – what is the added value? A review of alternative strategies for weighing harms and benefits as part of the assessment of animal research. *Laboratory Animals*, 53(1): 17-27.

The definitive version is available at

<https://doi.org/10.1177/0023677218783004>

not mutually exclusive and may be combined to make the most of their advantages while reducing the disadvantages of how harm benefit analysis is typically undertaken.

Introduction

The use of animals for research gives rise to a dilemma between the protection of the interests of the animals used and the promotion of the interests of those, mainly humans, benefitting from the research. To come to grips with this dilemma may in one respect be considered as an individual task in that the solution may depend on a moral outlook which can vary between individuals. One person may hold an anthropocentric view according to which human interests should always prevail, another person may hold an animal rights view according to which it is never right to impose harm on an animal to achieve benefits for humans, and a third person may hold a middle position that will allow the imposition of harms to animals if they give rise to proportionate human benefits. Research documents that there indeed are such differences in individual opinion¹. Although individual opinions on the moral limits of the use of animals for experimentation differ widely, it remains necessary to find solutions to set standards governing what scientists are permitted to do when it comes to animal use.

Such solutions have been implemented in different parts of the world through regulation – initially, mainly at the national level, but to an increasing extent through international initiatives including both formal measures, such as European Union (EU) legislation, and informal “soft law” norms governing the activities of international journals and the like.

Regulation of animal experimentation cuts through some of the moral disagreement. It typically does so by setting up some conditions that must be satisfied before the use of animals for research is permitted. The main conditions pressed into service so far have been, firstly, that the use serves a socially acceptable purpose, and secondly, that all possible efforts have been made to comply with the 3Rs of Replacement, Reduction and Refinement. To oversee this, across all developed countries systems have been set up where potential users of animals must apply for a permit which is issued only if an evaluation (usually undertaken by a committee comprising different competences and perspectives)² shows that the necessary conditions are met.

Increasingly, a third condition has been discussed: the idea that research should only be allowed if the expected benefits of the research (the outcome in terms of knowledge, improved treatment options etc. that can reasonably be expected to be generated) exceed the predicted harm to the animals. The idea for a more systematic approach to this was originally developed in a seminal paper³ in which Patrick Bateson argued that harms imposed on animals must be proportionate to the scientific value of the experiment. Bateson introduced the decision cube as a way to work out the acceptable level of animal suffering in an experiment relative to the value it has to humans, where the latter is defined in the two dimensions of “importance of research” and “likelihood of benefit”.

The need for harm-benefit analysis (HBA) of some kind was recognised when the first animal ethics committees were established by Swedish legislation in 1979⁴ and this kind of calculation became part of the assessment regime introduced by the 1986 Animals (Scientific Procedures) Act in the UK. The concept of HBA is now more or less embedded as common practice globally⁵, and as an explicit

requirement in a series of regulatory documents across the world. In the Guide for the Care and Use of Laboratory Animals⁶, which regulates publicly funded research in the United States, committees are “obliged to weigh the objectives of the study against potential animal welfare concerns”. In their international accreditation of institutions, the assessment body AAALAC International requires that

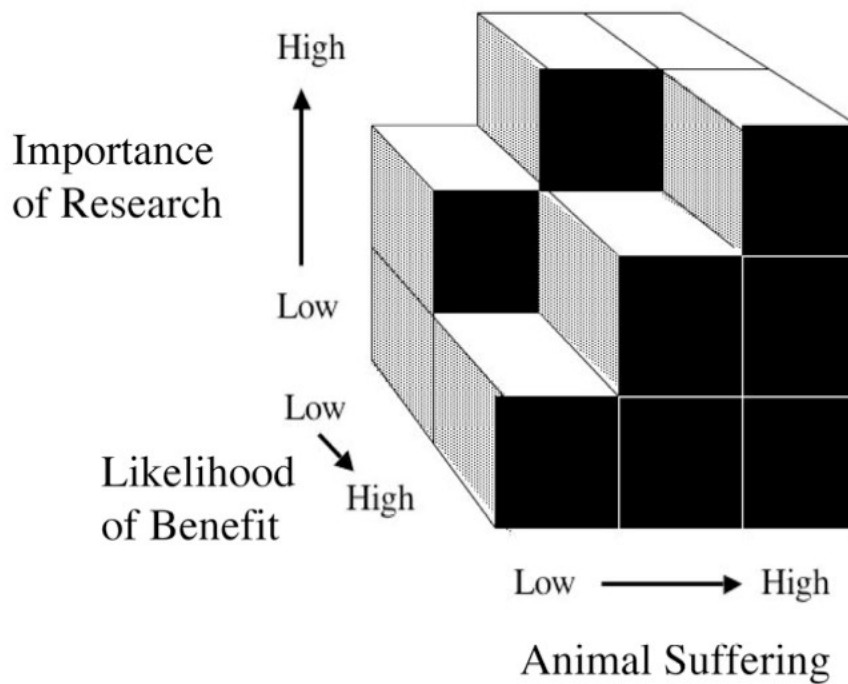


Figure 1. The Bateson cube for assessment of animal experiments⁴ (reprinted with permission from the Bateson family and Elsevier).

committees “as part of the protocol review process, will weigh the potential adverse effects of the study against the potential benefits that are likely to accrue as a result of the research”⁷. Directive 2010/63/EU, which forms the basis of the regulation of animal experimentation in all EU member states, requires that the project evaluation includes “a harm-benefit analysis of the project, to assess whether the harm to the animals in terms of suffering, pain and distress is justified by the expected outcome taking into account ethical considerations, and may ultimately benefit human beings, animals or the environment”⁸.

Typically, the HBA requirement is understood to be part of the ethics review which, in most places, is undertaken by committees of experts, stakeholders and in some cases also lay people. These committees discuss and reflect on the value of particular projects, taking into account the harm done to animals. Their approach is based on what we will term the “discourse model”.

However, over recent years, a number of authors have proposed more specific methodologies which, making use of rigorous criteria that are usually based on algorithms or graphic representations, operationalise the HBA in what we will term “metric models”. Although metric models have never, as far as we know, been implemented in animal ethics review in practice, they have triggered a lively debate on how to carry out an effective and transparent harm-benefit analysis (e.g.^{9, 10-12}). There also seems to be a growing interest in models which include a more rigorous method of HBA. Thus a recent US-European joint working group has proposed a model in which the different aspects of harm and benefit are colour-coded, with more intense colour indicating greater harm/benefit, as a way of producing an overview prior to the final assessment of ethical acceptability¹³.

Judging by the debate so far, it is clear that while there is a perceived need for more rigorous methods of harm-benefit analysis, there is also scepticism among practitioners about the methods that have been proposed. In this paper, we aim to address the ideas underlying the discussion and disagreements. We will argue that in essence there are two very different suggestions about how to organise the HBA. One is the discourse model, which is based on a dynamic and relatively loosely structured discussion within a multidisciplinary group of people. The other is the metric model, which prescribes a detailed system for attributing value to harms and benefits and, on that basis, delivers clear positive or negative verdicts.

In the next two sections, we outline the two models and ask how well each meets the different ideals of HBA. Against this background we turn, in the last section, to the question of what a satisfactory form of HBA would look like. We suggest it would need to combine elements from both models.

HBA as an integral part of ethical review in committee: the discourse model

The standard way of carrying out HBA has been set by the practice of ethics review committees. Here one of the elements included in Bateson’s model, “likelihood of benefit”, seems to play no or a very limited role. Although the original paper by Bateson could be interpreted as proposing a simple metric that did not require dialogue and moral deliberation, a later paper¹⁴ makes it clear that this was not what Bateson had in mind. His view seems rather to be that the application of his ideas will require deliberation in a diverse group of people: a process where “a lot of fair-minded people, who often start with utterly different views, are finding ways of reaching agreement”¹⁴.

The theoretical underpinnings of this approach are not terribly well worked out, but it is reasonable to assume that something like the following is the fundamental idea: by setting up committees on which not only experts, but also outsiders to animal experimentation such as representatives of animal protection groups and so-called lay members sit, it is possible to make room for dialogue where the committee members jointly, on behalf of society, draw boundaries around what is considered an acceptable research use of animals within the above-mentioned general framework. In other words, the public debate over the pros and cons of animal research is meant to crystallise in such committees.

The legal scholar Katarina Alexius Borgström⁴ has presented evidence from official documents that the Swedish system, with its committees of equal numbers of people linked to animal experimentation and lay people, was indeed set up to allow deliberation across initially conflicting views and interests

so that decisions that could be viewed as genuinely joint decisions serving to balance different individual moral views are arrived at.

This could, in turn, be underpinned by various forms of ethical or political thinking. That thinking could range from so-called procedural or discourse ethics¹⁵, according to which moral views should be validated through a social dialogue, across the view, found in the later works of John Rawls¹⁶, according to which moral disagreements in a social setting should, as far as possible, be decided by means of so-called overlapping consensus, and on to more pragmatic views favouring moral compromise¹⁷.

In light of this, the methodological ideal for HBA could be summarised in something like the following way. It is always necessary to check first whether a project satisfies the requirements that it is being conducted for an appropriate purpose and that every practically possible effort has been made to apply the 3Rs. Also, the expected harm and benefit must be estimated. These tasks can be handled by legal or scientific experts. Following this, the committee in charge of reaching a verdict on whether a project should be accepted can engage in deliberation in terms of HBA. Here the committee will need to engage in a dialogue about whether the expected benefit is of a kind and size that can justify the anticipated harms to the laboratory animals. When doing this, the committee must, to ensure fairness in the assessment of other projects, have a view to a more general policy level – looking across projects and ideally also across committees who work within the same legal remits. This means that proper harm benefit assessment requires a double perspective including both the specific project and the precedent the verdict sets. The general idea is that, over time, committees will develop a nuanced picture of where to place the bar determining how much harm to laboratory animals is acceptable for the sake of achieving various forms of benefit.

This account of the workings of HBA within the framework of ethics review in committees may be challenged on both theoretical and practical grounds. From a theoretical point of view, it can be argued that the idea of a sensible middle ground where disagreements about animal use can be argued out between fair-minded people who are not in the grip of “bigotry and fanaticism”¹⁴ is an attempt to impose a specific moral view on the whole of society. Those who favour an animal rights view according to which animal experimentation should be banned will claim that they are not irrational fanatics or bigots – they just appeal to reasons that differ from those accepted by the majority. Instead of beefing up the majority view as the product of an imaginary social compact, HBA could instead be described as a process allowing the majority moral view overruling the view of the minority.

From another perspective, a system is only as good as it works in practice. How well the real functioning of committees deliberating over animal experiments corresponds to the imagined ideal has been addressed empirically in several studies. A common observation focuses on the influence of scientists and the associated concern that non-scientist voices are not heard. This imbalance may be a result of the sheer number of scientists. Studies of committee composition in both the US and Canada show that scientists are present in much greater numbers than any other group^{18, 19}.

Average contributions of scientific and non-scientific members of committees have also been shown to differ. Silverman and colleagues²⁰ analysed the deliberations of 10 committees operating in top-funded US biomedical research institutions. They found that the scientists contributed 50% more

comments than the community members – indeed, taking into account the higher total number of scientists, they calculated that overall scientists contributed nearly 10 times as much as community members to the discussion. Examining review procedures in three Canadian research institutions, Houde et al.²¹ found that discussion was dominated by the committee chair, the two committee members who were animal researchers and the veterinarian committee member. Only 33% of contributions were made by the remaining six members (non-animal researcher member, two technicians, lay member, administrative officer and student member). In interviews, Canadian committee members also commented on numerical representativeness, raising the “concern that community members would carry little weight in discussion because they were outnumbered by affiliated and mostly scientist members”¹⁸. Ideland²² argues that even in Sweden, where committees are composed with equal numbers of members from the scientific community and lay members, scientists have power over the agenda. Their expertise in science, which is both an important knowledge domain for the committee discussion and generally inaccessible to non-scientists, is likely to give them an advantage independently of whether they are in majority. Scientists can also come to dominate discussion by taking on certain roles. Silverman and colleagues²⁰ investigated the relationship between the initial presentation of a protocol by a committee member and the subsequent committee discussion. The topics introduced by the presenter were found to be those which dominated the committee discussion. Given that the majority of presenters were scientists, this was considered an additional factor in the scientists’ domination of the discussion.

Does an imbalance in committee composition and discussion affect the outcome of the reviews in general and the harm-benefit analysis in particular? The fact that most applications are in fact approved, at least after revision, has been interpreted as an approval bias resulting from committees where a large majority of members have a vested interest in research^{23, 24}. Background also affects the types of comment different committee members make in the discussion. Silverman et al.²⁰ found that community members were 1.7 times more likely than the scientists to discuss pain and distress.

What do the committees actually discuss, and do they perform a harm-benefit analysis? In her interviews with 20 Swedish committee members, Ideland²² found that the majority of interviewees “agree that the animal ethics committees are discussing the ‘wrong’ questions” and do not engage in analysing harm versus benefit. A suggested explanation, in addition to the scientific discourse dominating the discussion, is that a technical discussion allows the committee to reach consensus, and that consensus would probably be unattainable for a harm-benefit analysis. Schuppli²⁵ interviewed 28 Canadian committee members and found that half of them did some kind of HBA, that they found it challenging, and that their approaches varied. Only a quarter of the interviewees generally required benefit to outweigh harms. Several highlighted the impossibility of conducting HBA because the benefits were too difficult to assess. Many added that there must be an upper limit for harms irrespective of benefits.

A final important question concerns the consistency and potential standardisation of animal ethics review. Plous and Herzog²⁶ studied 50 institutional committees in the US. Each committee provided recently assessed protocols for a second assessment by another committee. They found no significant relation between the first (original) and the second (experimental) recommendation. In the cases where the first and second reviews made different recommendations, the second committee was

nearly always more negative than the original review. The only aspect over which there was a strong correlation between the first and the second review was pain assessment, which was also the only aspect for which the committees were provided with a detailed scale with examples.

The conclusion to be drawn, looking at the discourse model both from a theoretical and a practical perspective, is that there is a problem: there is lack of consistency across decisions taken by different committees, and it is difficult to ensure that common standards are applied in a transparent way. The metric models can clearly be seen as an answer to this problem.

Metric models of HBA: quantifying ethical deliberation

The promise of metric models is to facilitate the HBA and allow for a transparent, consistent and standardised decision-making process. In its most thoroughgoing version, the ideal is a methodology for HBA which can be used even without the help of a review committee. With information provided by the applicant, harms and benefits can be assigned and weighed using a predefined algorithm.

The metrics model idea is not new: Canadian biomedical physicist David G. Porter was a first mover in this approach when publishing his “scoring system” in *Nature* in 1992²⁷. In this system eight domains are scored and aggregated: a) aim of experiment, b) realistic potential of experiment to achieve objective, c) species of animal, d) pain likely to be involved, e) duration of discomfort or distress, f) duration of experiment, g) number of animals, h) quality of animal care. In each domain a maximum of 5 points can be achieved. A low total score suggests that the project is less problematic than one with a higher total score. The idea of a scoring system using predefined criteria as a measure of HBA and applying cut-off scores has been taken up by a number of other scholars (e.g. ²⁸⁻³³) who have proposed different models based on essentially the same idea.

The overall theoretical idea of capturing the complexity of HBA in a systematic tool is not entirely different from the ideal of the discourse model and the work of committees. In both cases, relevant criteria need to be identified, their degree of satisfaction has to be assessed, and a conclusion needs to be reached in which all factors are considered. However, whereas in the discourse model the components of HBA are often only implicit, HBA components must be explicitly identified and in place beforehand in metric models. The four most important components are the following: a) a defined set of criteria which comprise the harm and benefit dimensions to be included; b) the relative weights of the individual criteria; c) operational factors to identify and measure how well each criterion is fulfilled; d) a procedure that aggregates total harms and benefits into a final HBA outcome. If these four components can be adequately established, metric models will allow for a standardised and transparent weighing and decision-making process. This in turn will enable consistent standards to be applied across projects. It will also allow applicants to know beforehand which criteria are of relevance and what kinds of information have to be provided for the HBA to be carried out in the project application. By applying the same metric methodology as the authority, an applicant would know in advance whether the HBA is likely to result in a positive or negative outcome. Potentially, it seems, the metric methodology would contribute massively to transparency, reduce identified inconsistencies in the evaluation process³⁴, lead to greater harmonisation at the national and international level, and improve efficiency.

Thus, in its ideal version, the metric model seems very promising. However as already mentioned, criticisms have been made of various aspects of previously proposed models (for an overview cf. ⁹, ¹⁰), and the overall idea of metric HBA faces a number of theoretical and practical challenges.

The primary theoretical challenge lies in how best to capture ethical deliberation in a metric model. Brønstad and co-authors⁹ claim that “moral dilemmas cannot/shall not be solved by arithmetics.” Essentially, the point here made is that a measuring and weighing system *eliminates responsible deliberation and judgement*, an issue also raised by e.g. Bout et al.³⁵.

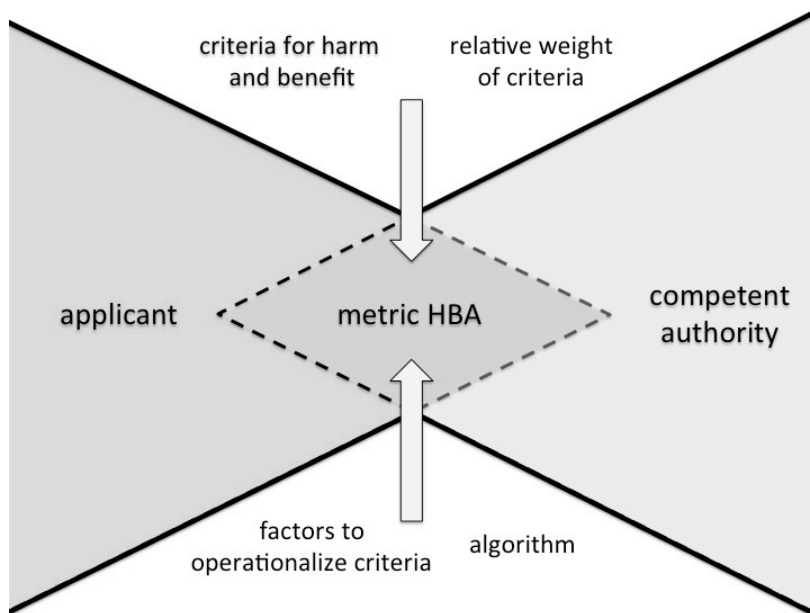


Figure 2. Components of the metric harm–benefit analysis (HBA) that systematically guide application and evaluation.

Although it is true that if applied as the only evaluation instrument, the metric model replaces practical deliberation in dialogue, it is not true that ethical thinking does not play a role in metric models. The ethical thinking takes place, not when the evaluation instrument is used, but when it is being developed. Numerous ethical and legal judgements are needed to assess whether particular factors are of relevance, and to attach suitable weightings to these factors, and these need to be integrated as criteria in the resulting standardised metric model.

However, the fact that ethical judgements are made primarily in the development of the model and not in its application is problematic in several ways. One is the question of accountability. Project evaluation decisions have legal implications – the evaluation is a mechanism used to implement the legislation that regulates the use of animals in experiments (e.g. ²). As such, somebody needs to be accountable for the resulting decisions. When they are the outcome of ethical reasoning by a

committee that has the mandate to deliberate over project applications, accountability lies with that committee. But with a metric model, ethical judgements are made when criteria are defined and weighted in the development of the method. It is much less clear who is accountable for decisions that are the outcome of predefined metrics. The assumptions that go into these metrics are in many cases not spelled out – and more importantly, they are not deliberated upon and accepted by those who have a mandate to do so.

The accountability issue can be dealt with in part by ensuring that deliberation takes place in the model development and is done by people who are mandated by society to do so. This could be through a participatory process where stakeholders (applicants, competent authorities, industry, animal protection groups, 3R specialists) are represented and agree on the criteria, their relative weight, the factors identifying the degree of fulfilment, and the algorithm. However, such participatory processes are highly demanding and, against a background of polarized debate over animal use, at risk of failure.

Indeed, experience from an Austrian project³⁶ supports the worry that participatory processes will not lead to consensus. The aim of this project was to develop a standardised method for the evaluation of animal research proposals, and HBA in particular, on the basis of objective criteria. The resulting method was supposed to be used in their project evaluations by the competent authorities *without* the help of an ethics committee. Hence, the aim was to develop an algorithm which generates the result of the HBA on the basis of information provided by the applicant “automatically.” Further, the intention was to develop and evaluate the method in a participatory process that included stakeholders from the industry, representatives of science and research institutions, the relevant authorities, representatives from NGOs and 3R specialists. However, in none of the above-mentioned four basic components of HBA did the Austrian project reach consensus among stakeholders and experts groups. An important source of disagreement was the idea that a single set of criteria could be applied across all fields of research. The project showed that attempts to press the complexity of the various research fields into a single set of predefined criteria are unlikely to find shared support.

In addition to the challenges inherent in reaching consensus over a politically charged issue, the HBA method also needs to overcome a difficult methodological issue. Since the algorithm in a metric model has to accommodate weightings of different value dimensions and reduce them to one “currency”, easy solutions for the weighing procedure are not very likely^{32, 36}. The problem of adding up “apples and oranges” comes to the fore in metric methodologies¹².

A second methodological problem has to do with the complex and varying nature of projects. Contextual factors which only become evident as one is looking at a particular research project can be handled in case-by-case deliberation in a more dynamic system. In contrast, a universal methodology, applicable across all of the different kinds of project that may have to be evaluated in a given country, means that every possible scenario and problem has to be anticipated and dealt with at the design and development stage. This is unlikely to be successful. Instead a “one size fits all” approach will be forced upon projects for which it is not necessarily well suited.

A concrete example mentioned by a number of authors is that existing models bias HBA in favour of applied research (e.g. ³⁷⁻⁴²) by asking for expected benefits to be measured in concrete quantitative

terms such as number of patients affected. Indeed, the uncertainties connected with the prospective assessment of expected benefits are a major concern for this kind of evaluation⁴³. With good reason, Brønstad and colleagues categorise benefits as “future promise” which may never be realised⁹ – a categorisation suggesting that the ambition of quantifying expected benefits through any kind of numeric measure is questionable.

Innumerable scenarios and issues can be imagined which tend to confirm that the outcome of a project (knowledge) might not be turned into quantifiable benefit³⁹. An ethics review committee with a mandate to assess expected harm and benefit can, and has to, deal with such scenarios and issues. In the metric model the issues become a problem, since they have to be predefined in terms of operational factors.

The metric model has been criticised for its commitment to the idea of a clear-cut methodology in a field where clear-cut solutions are considered inappropriate or illusory^{5, 29, 44, 45}. The need for explicit deliberation and the ability to consider contextual factors are consistent with these criticisms. The problems, both of accountability and the lack of consideration of project-specific contextual factors, are aggravated if the metric is applied wholly outside the context of wider ethics review by committee. Although in practice this happens only rarely, it is in fact how HBA was intended to function in the evaluation of animal experimentation in Austria³⁶.

In summary, the metric methodologies promise a structured, transparent and standardised process of HBA. But although they provide a shared basis for weighing harms and benefits for applicants and the authority alike, their severe problems on the theoretical and practical level make it unlikely that a thoroughgoing metric model will make good on its promises if it is applied without additional support in a project evaluation.

How to achieve the added value of HBA? Combining the two models

The last two sections described and analysed important aspects of the two main methods of HBA in ethical review of animal experiments. The ideas behind, and the attractions and shortcomings of, the two models were explained. At present only the discourse method is used. Social science studies were summarised in an effort to understand how the ideals of HBA are secured with that method in practice. An overview of the ideals and challenges connected with each of the two methods is presented in Table 1.

Table 1. Overview of the ideals and the challenges of the discourse and metric models of harm-benefit analysis

	Discourse Model	Metric model
Ideals	<ul style="list-style-type: none"> ▪ Ethical judgements should be validated through a social dialogue ▪ Disagreements about animal experimentation can be argued out between fair-minded people ▪ Committees over time develop standards for acceptable harm-benefit balance 	<ul style="list-style-type: none"> ▪ Transparency of criteria and consistent weighing ▪ Standardisation and harmonisation across projects, and evaluation procedures applied nationally and internationally ▪ Clear-cut predictable decision-making
Challenges		
<i>Accountability</i>	As ethical judgements take place within committee deliberation over individual projects => accountability for a decision lies with the committee who makes it	As ethical judgement takes place when criteria are defined and weighted in the development of the method => unclear who is accountable for decisions made which are the outcome of predefined metrics
<i>Developing and operating the method</i>	<p>To establish a method that allows the ideal to be met it is necessary to find a way to structure the discussion that allows all voices to be heard, that keeps a relevant focus, and that operates in a consistent manner</p> <p>In practice:</p> <ul style="list-style-type: none"> ▪ A thoroughly developed methodology is often missing ▪ Social and professional roles strongly influence the deliberation and the composition of committees shapes the outcome ▪ Unguided discussion results in focus on issues other than HBA ▪ Limitations in the standardisation of procedure and outcome ▪ Non-transparent criteria 	<p>To establish a method that allows the ideal to be met it is necessary to work out the complete and complex framework for a metric HBA:</p> <ul style="list-style-type: none"> ▪ defining a set of criteria covering the harm and benefit dimensions; ▪ identifying the relative weights of individual criteria; ▪ finding operational factors to measure how well criteria are fulfilled; ▪ formulating a procedure that computes total harms and benefits and generates a final outcome <p>In practice:</p> <ul style="list-style-type: none"> ▪ Unlikely to achieve consensus of stakeholders and experts on the composition of the methodology
<i>Dealing with the diversity of projects</i>	<ul style="list-style-type: none"> ▪ Contextual factors can be handled in case-by-case deliberation 	<ul style="list-style-type: none"> ▪ “One size fits all mentality” of the particular metric underestimates the complexity and diversity of projects ▪ Risk of biased results and inappropriate outcomes of the HBA if a single predefined metric is applied to all types of project

It seems obvious that both models have merit, but face challenges, and that exclusive application of either one of them is unlikely to offer an ideal solution. In the following we will therefore discuss possible compromises. However, let us first take a step back and ask to what extent HBA is a significant addition to the project evaluation process.

Against the backdrop of what has been said, the following five observations help to put the added value of HBA in perspective. First, in whatever form, HBA raises awareness that reducing harm and applying the 3Rs in animal research is no longer sufficient. Focus on the benefit dimension marks a change in the mind-set around animal research. Second, the demand that we balance harm to animals and benefits to others, primarily human beings, emphasises that animals have to be considered in a significant way. Third, HBA has given rise to valuable discussions about how decisions are taken in review committees and a call for transparency in approval procedures. Fourth, and connecting with transparency, HBA helps to underline the need for consistency across projects, committees and countries, which is of crucial importance for harmonisation. Fifthly, the idea that applicants should receive fair treatment (in among other things legal terms), and that at the same time the needs of research animals should be given fair consideration, is preserved in HBA.

Hence, potentially, there is significant added value to be had by incorporating HBA into the review process. The question is whether the unresolved challenges facing the discourse model and the metrics model can be overcome by combining them.

The core idea of the metric model is promising: applicants and the authorities work on the basis of a shared set of criteria which operationalise the harm and benefit dimension of particular experiments (Fig. 2). An explicit set of criteria allows not only for more transparency, consistency and fairness, but also for improved efficiency in the authorisation procedures of committees. Information on relevant criteria could easily be provided for the applicants and be picked up by committee members and authorities. Further, explicit criteria could facilitate a uniform development across committees and countries. The results on the various harm and benefit dimensions can be documented transparently, and step-by-step, in a way that facilitates retrospective analysis and allows committees to monitor their standards.

Of course, a set of criteria does not fully answer the question of how to weigh harms and benefits. In this regard a procedural approach in which solutions are located in a social dialogue, as is the ideal of the discourse model, can be used as a guiding idea. As we have seen, for various reasons it is not very likely that agreement on a particular algorithm (metric model) will be achieved prospectively. However, for a decision to be reached in committees, weighing has to take place anyhow (implicitly or explicitly) in one way or another.

The results of reviews undertaken in committees could be used to retrospectively learn about, and reflect on, plausible weightings and weighing strategies to be aligned with the given legal framework. This approach would make it possible to apply context-sensitive weighing procedures step-by-step. Prospectively, they could serve as starting points in committees.

In terms of *methodological challenges*, a merger of the two models could overcome the shortcomings of the discourse model – shortcomings seen predominantly in a bias towards science resulting from committee composition in combination with unstructured, inconsistent and non-transparent

procedures. An explicit set of criteria would focus and direct the work of review committees and help to avoid a bias, in their discussions, towards the values of committee members in dominant roles. Uncertainties will, of course, remain when harms and benefits are assessed prospectively. However, the committee would be mandated to decide or advise in conditions of uncertainty.

The combination of the two models would facilitate transparent dealing with uncertainty and a standardised documentation of the decision-making process. On a practical level, the merger would allow room for flexibility within a transparent framework. Since a “one size fits all” metric model with a predefined algorithm seems to be too ambitious, or even misconceived, the development of criteria which guide committees through the HBA and allow for reflection and debate seems desirable. This methodology ought to enable important aspects of HBA to be operationalised within a given framework without an agreed metric. The criticism of metric models over their inflexibility can be avoided, and the complexity of “non-standard” projects can be taken into account.

Therefore, we suggest that a combination of the two models offers the most promising way forward. When HBA is built on the strengths of the metric model, transparent, consistent, fair and efficient approval processes are attainable. The risk that the discourse model will be used as a smoke screen is also avoided and its strengths maintained. It allows for flexibility, contextual sensitivity, and the integration of expert knowledge. In practice, the final decision will be based on a dialogue in light of comprehensive input delivered by the metric model relating to criteria which comprise the harm and benefit dimensions. In sum, we believe that the full added value of HBA can be realised, if applicants, authorities and review committees work on the basis of explicit and transparent criteria that allow for fair-minded dialogue and transparent decisions.

References

1. Lund TB, Mørkbak MR, Lassen J, et al. Painful dilemmas: A study of the way the public's assessment of animal research balances costs to animals against human benefits. *Public Understanding of Science* 2014; 23: 428-444. DOI: 10.1177/0963662512451402.
2. Olsson IAS, Silva SPd, Townend D, et al. Protecting Animals and Enabling Research in the European Union: An Overview of Development and Implementation of Directive 2010/63/EU. *ILAR Journal* 2016; 57: 347-357. DOI: 10.1093/ilar/ilw029.
3. Bateson P. When to experiment on animals. *New Sci* 1986; 109: 30-32.
4. Alexius Borgström K. *Djuren, läkarna och lagen: en rättslig studie om djurförsöksetik*. Uppsala: Iustus Förlag, 2009.
5. Griffin G, MacArthur Clark J, Zurlo J, et al. Scientific uses of animals: harm-benefit analysis and complementary approaches to implementing the Three Rs. *Revue scientifique et technique (International Office of Epizootics)* 2014; 33: 265-272.
6. National Research Council of the National Academies. *Guide for the Care and Use of Laboratory Animals*. 8th ed. Washington, D.C.: The National Academies Press, 2011.

7. AAALAC international. https://www.aaalac.org/accreditation/faq_landing.cfm (2018, accessed Apr 13th, 2018).
8. European Union. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. *Official Journal of the European Union L276* 2010; 53: 33-79.
9. Brønstad A, Newcomer CE, Decelle T, et al. Current concepts of Harm-Benefit Analysis of Animal Experiments - Report from the AALAS-FELASA Working Group on Harm-Benefit Analysis - Part 1. *Lab Anim-Uk* 2016; 50: 1-20. DOI: 10.1177/0023677216642398.
10. Alzmann N. *Zur Beurteilung der ethischen Vertretbarkeit von Tierversuchen*. Tübingen: Narr Francke Attempto, 2016.
11. Borchers D and Luy J. *Der ethisch vertretbare Tierversuch. Kriterien und Grenzen*. Paderborn: Mentis, 2009.
12. Grimm H. Turning Apples into Oranges? The Harm–Benefit Analysis and How to Take Ethical Considerations into Account. *Atla-Altern Lab Anim* 2015; 43: 22-24.
13. Laber K, Newcomer CE, Decelle T, et al. Recommendations for Addressing Harm-Benefit Analysis and Implementation in Ethical Evaluation - Report from the AALAS-FELASA Working Group on Harm-Benefit Analysis - Part 2. *Lab Anim-Uk* 2016; 50: 21-42. DOI: 10.1177/0023677216642397.
14. Bateson P. Ethics and Behavioral Biology. *Advances in the Study of Behavior* 2005; 35: 211-233. DOI: 10.1016/S0065-3454(05)35005-4.
15. Habermas J. *Moral Consciousness and Communicative Action*. Cambridge: MIT press, 1990.
16. Rawls J. *Political Liberalism*. New York: Columbia University Press, 1993.
17. Kupper F and De Cock Buning T. Deliberating Animal Values: a Pragmatic—Pluralistic Approach to Animal Ethics. *Journal of Agricultural and Environmental Ethics* 2011; 24: 431–450.
18. Schuppli CA and Fraser D. Factors influencing the effectiveness of research ethics committees. *J Med Ethics* 2007; 33: 294-301. DOI: 10.1136/jme.2005.015057.
19. Hansen LA, Goodman JR and Chandna A. Analysis of Animal Research Ethics Committee Membership at American Institutions. *Animals (Basel)* 2012; 2: 68-75. DOI: 10.3390/ani2010068.
20. Silverman J, Lidz CW, Clayfield J, et al. Factors Influencing IACUC Decision Making: Who Leads the Discussions? *Journal of Empirical Research on Human Research Ethics* 2017; 12: 209-216. DOI: 10.1177/1556264617717827.
21. Houde L, Dumas C and Leroux Trs. Animal Ethical Evaluation: An Observational Study of Canadian IACUCs. *ETHICS & BEHAVIOR* 2003; 13: 333-350. DOI: 10.1207/S15327019EB1304_2.
22. Ideland M. Different views on ethics: how animal ethics is situated in a committee culture. *J Med Ethics* 2009; 35: 258-261. DOI: 10.1136/jme.2008.026989.

23. Kolar R and Ruhdel I. A Survey Concerning the Work of Ethics Committees and Licensing Authorities for Animal Experiments in Germany. *Altex-Altern Tierexp* 2007; 24: 326-334. DOI: DOI 10.14573/altex.2007.4.326.
24. Hansen LA. Institution animal care and use committees need greater ethical diversity. *J Med Ethics* 2013; 39: 188-190. DOI: 10.1136/medethics-2012-100982.
25. Schuppli CA. Decisions about the Use of Animals in Research: Ethical Reflection by Animal Ethics Committee Members. *Anthrozoös* 2011; 24: 409–425.
26. Plous S and Herzog H. Reliability of Protocol Reviews for Animal Research. *Science* 2001; 293: 608-609. DOI: 10.1126/science.1061621.
27. Porter DG. Ethical scores for animal experiments. *Nature* 1992; 356: 101-102. DOI: 10.1038/356101a0.
28. de Cock Buning T and Theune E. A Comparison of Three Models for Ethical Evaluation of Proposed Animal Experiments. *Animal Welfare* 1994; 3: 107-128.
29. Scharmann W and Teutsch GM. Zur ethischen Abwägung von Tierversuchen. *ALTEX* 1994; 11: 191-198.
30. Mand U. Über die in § 7 Abs. 3 des Tierschutzgesetzes geforderte Abwägung ethischer Vertretbarkeit von Tierversuchen. *Der Tierschutzbeauftragte* 1995; 3: 229-234.
31. Mand U. *Zur Abwägung ethischer Vertretbarkeit von Tierversuchen gemäß § 7 Abs. 3 des derzeit geltenden Tierschutzgesetzes (TierSchG), dargestellt am Modell Sepsisforschung am wachen Schwein*. Diss. med. vet., Univ. Gießen, Gießen, 1995.
32. Stafleu FR, Tramper R, Vorstenbosch J, et al. The ethical acceptability of animal experiments: a proposal for a system to support decision-making. *Lab Anim-Uk* 1999; 33: 295-303. DOI: 10.1258/002367799780578255.
33. SAMW/SCNAT SAoMSaSAoS. Ethische Güterabwägung bei Tierversuchen. Eine Vorlage für die Selbstprüfung, <http://tki.samw.ch/> (2017, accessed Dec 20th, 2017).
34. Varga O. Critical Analysis of Assessment Studies of the Animal Ethics Review Process. *Animals (Basel)* 2013; 3: 907-922. DOI: 10.3390/ani3030907.
35. Bout HJ, Fentener van Vlissingen JM and Karssing ED. Evaluating the ethical acceptability of animal research. *Lab Anim-Uk* 2014; 43: 411-414. DOI: 10.1038/labam.572.
36. ACC TACoCP. <http://www.vetmeduni.ac.at/messerli/forschung/forschung-ethik/projekte/kriterienkatalog-zur-bewertung-von-tierversuchsantraegen/> (2017, accessed Jan 10th, 2018).
37. Abbott A. Basel Declaration defends animal research: Dialogue with public is key to reducing opposition over the use of lab animals. *Nature* 2010; 468: 742.

38. Graham ML and Prescott MJ. The multifactorial role of the 3Rs in shifting the harm-benefit analysis in animal models of disease. *European Journal of Pharmacology* 2015; 759: 19-29. DOI: 10.1016/j.ejphar.2015.03.040.
39. Grimm H and Eggel M. White Paper and Colourful Language: Toward a Realistic View of Animal Research. *Atla-Altern Lab Anim* 2017; 45: 101-103.
40. Peggs K. Nonhuman Animal Experiments in the European Community: Human Values and Rational Choice. *Society & Animals* 2010; 18: 1-20.
41. Pfister R, Gyger M, Bugnon P, et al. Editorial ethical guidelines: what about the evaluation of the benefits to humans? *Lab Anim-Uk* 2012; 46: 176. DOI: 10.1258/la.2012.011129.
42. Schiermeier Q. German authority halts primate work. *Nature* 2008; 455: 1159. DOI: 10.1038/4551159a.
43. Grimm H, Eggel M, Deplazes-Zemp A, et al. The Road to Hell Is Paved with Good Intentions: Why Harm–Benefit Analysis and Its Emphasis on Practical Benefit Jeopardizes the Credibility of Research. *Animals (Basel)* 2017; 7: 70. DOI: 10.3390/ani7090070.
44. Mellor DJ. Comprehensive Assessment of Harms Caused by Experimental, Teaching and Testing Procedures on Live Animals. *Atla-Altern Lab Anim* 2004; 32: 453-457.
45. Voipio H-M, Kaliste E, Hirsjärvi P, et al. Nordic-European Workshop on Ethical Evaluation of Animal Experiments. Workshop Report on the Cost-Benefit Principle in Hanasaari, Helsinki, Finland 7-9 November 2003. *Scandinavian Journal of Laboratory Animal Science* 2004; 31: 251–267.